Formalization of the input/output retinal transformation regarding non-standard ganglion cells behavior

Elaa Teftef, Thierry Viéville

Formalization of the input/output retinal transformation regarding non-standard ganglion cells behavior

E. Teftef, T. Viéville*

Inria Mnemosyne `http://team.inria.fr/mnemosyne`, France.

September 28, 2012

The aim of the present work is the realization of a simulation software of mesoscopic biological models of non-standard features of the retina, i.e. K-cells (results are very different for X and Y cells). This builds a link between biological models and non-trivial image processing algorithms of computer vision. The software created here simulates some functionality of the retina at the network scale (not at the cell level), considering real natural images and without being adjusted on each cell.

The models is based on three major biological assumptions, discussed in (Teftef et al., 2012; Teftef, 2012), derived from recent contributions regarding "the evidence that the early steps of mammalian vision are more diverse and more interesting than is usually imagined, so that our understanding of the later stages is in trouble right from the start" (Masland & Martin, 2007), i.e. the fact that "the retina solves a diverse set of specific tasks and provides the results explicitly to downstream brain areas", including "sophisticated spatio and temporal pattern recognition" (Schwartz & Michael, 2008) or "segregation of object and background motion" (Olveczky, Baccus, & Meister, 2003).

On one hand, we consider that what the retina is able to compute can be interpreted in terms of sophisticated second-order visual cues, including spectrum amplitude signature, since this is equivalent to linear/non-linear filtering, while it has been observed that it is a relevant parameter to characterize natural image categories (Torralba & Oliva, 2003), see Fig 2. On the other hand, we interpret

*Supported by the CONICYT/ANR KEOpS project and CORTINA associated team.

the non-local filtering of the different visual streams as a mechanism of image segmentation, allowing the optical nerve to perform a huge data compression tuned to what is specific in term of visual data. As a consequence, both fine scale and large scale visual information is output. The bio-plausible implementation of such mechanism has been extensively discussed elsewhere (Viéville, Chemla, & Kornprobst, 2007), see Fig 3 for results. Finally, we consider that the retina is able to detect static or time-varying visual events, and propose after (Viéville & Crahay, 2004) an biologically plausible implementaton of such mechanism, evaluated on a realistic data set, see Fig 4.

We thus propose here to use a variational framework to model and simulate, given natural image sequences, the mesoscopic collective non-standard behavior of some retinal input/output functions that correspond to the output of a sub-class of the so-called konio-cells, as sketched out in Fig. 1. See (Teftef et al., 2012) for an extended discussion of the class of retinal cells addressed by such a modeling. We hypothesize that from sophisticated temporal pattern recognition, to image segmentation, or specific natural statistical recognition, a unique generic two-layers non-linear filtering mechanism with feedback is implemented in the biological tissues, while not the individual but the collective behavior of the retinal cells answer for such input/output functions. Taking the retinal architecture and related biological constraints into account and considering the wider class of early-vision non-standard sensory-motor functionality known as non-standard behaviors, we use computer vision methods to propose an effective link between the observed functions and their possible implementation in the retinal network. With such framework, it is not possible to assume that the brain visual system stands on the reception of an homogeneous visual pipe of filtered images from the retina, whereas it is connected to several heterogeneous sources of information at different spatial and visual scales, and different integration level, while tuned to the natural image statistics. Roughly speaking, this encourages bio-inspired systems to compute in parallel several information streams depending on the sensory-motor task, instead of thinking of large information pipe. Furthermore, visual prostheses are likely to be though as a plural of sensors with a non-negligible data processing, before feeding the nervous system.

References

Masland, R. H., & Martin, P. R. (2007, August 07). The unsolved mystery of vision. *Current Biology*, *17*(15), R577–R582.

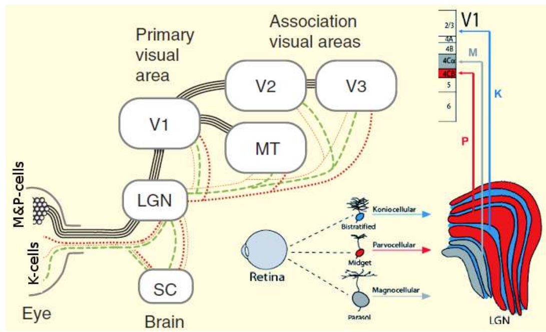

Figure 1: Schematic representation of projections of the K-cells considered here, after (Masland & Martin, 2007) and (Nassi & Callaway, 2009). K-cells represent 10% of the retinal projections, with large (about 10deg) visual fields and react to visual events. See (Teftef et al., 2012) for details.

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.*, *10*(5), 360–372.

Olveczky, B. P., Baccus, S. A., & Meister, M. (2003, May 22). Segregation of object and background motion in the retina. *Nature*, *423*(6938), 401–408.

Schwartz, G., & Michael. (2008, April 01). Sophisticated Temporal Pattern Recognition in Retinal Ganglion Cells. *Journal of Neurophysiology*, *99*(4), 1787–1798.

Teftef, E. (2012). *Formalisation de la transformation analogique / événementielle des mécanismes non-standards des cellules ganglionnaires de la rétine*. Unpublished master's thesis, Université de Paris 6 et Université El Manar de Tunis.

Teftef, E., Alexandre, F., Carvajal, C., Cessac, B., Escoba, M.-J., Palacios, A., et al. (2012). Modeling non-standard retinal in/out function using computer vision variational methods. *J. Physiol. (Paris)*. (submitted)

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. In *Network: Computation in neural systems* (Vol. 14, pp. 391–412).

Viéville, T., Chemla, S., & Kornprobst, P. (2007). How do high-level specifications of the brain relate to variational approaches? *J. Physiol. Paris*, *101*.

Viéville, T., & Crahay, S. (2004). Using an hebbian learning rule for multi-class svm classifiers. *Journal of Computational Neuroscience*, *17*(3), 271–287.

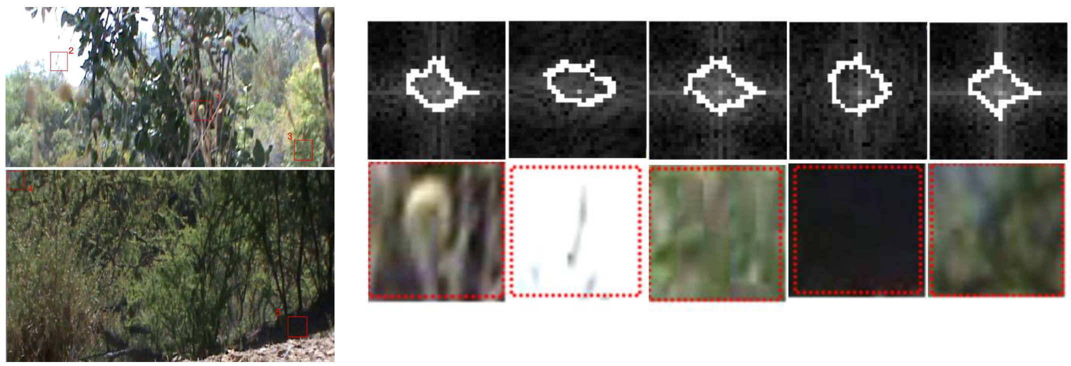

Figure 2: Considering a multi-stream computational framework, including spectrum amplitude signature. For five different zones of natural images labeled on the left picture, the spectrum amplitude and the 50% level curves are drawn. From left to right : 1. a flower, 2. a piece of sky with an artifact, 3. a tree zone, 4. a dark zone, and 5. a third tree zone perturbed by a piece of sky. Interesting enough is the fact that zones 3. and 5., despite their very different visual aspect have a rather similar spectrum signature, different from 1. Though zones 2. and 4. have spurious spectral responses (for 2. this is the artifact in the sky that yields the spectrum and for 4. the spectrum is not relevant in such a dark zone), since the system is also using the local color as a cue, this yields no mistake. We numerically observed that not only the whole image second-order statistics (Torralba & Oliva, 2003), but also very local second-order statistics seem to be relevant for natural image categorization. More precisely, in the present numerical implementation limited to static images, the color and the sum of the spectrum amplitude responses at 50% of the maximal amplitude are taken as cues to characterize the local texture, in order to perform region detection. The fact that such very simple scalar parameter calculated on the spectrum magnitude leads to relevant results (see Fig. 4) is a good argument in favor of the proposed choice. The generalization to spatio-temporal volumes, thus 2D+T spectra, is straightforward and is a direct perspective of this work.

Figure 3: Qualitative examples of non-local filtering using diffusion mechanisms as discussed in (Viéville et al., 2007) regarding their biological implementation. Lower images are the non-local filtering output of one channel (here intensity) computed on the corresponding upper image. Interesting enough, several key points can be observed here : The output is morally an “artificial image” corresponding to the “natural” input, where the forms in the image have been preserved. More precisely, the edges (e.g., the flowers boundaries) are preserved even when the image is “smoothed” : The fact the mechanism filters the noise and not the edges, is simply due to the non-linearity of the diffusion operator, since small variations diffuse and are thus filtered, whereas higher intensity variations related to edges do not. Furthermore, diffusion is anisotropic, more precisely performed along the edge, but not across it (in a more mathematical language : Tangential to the local edge orientation, but not orthogonal to it), thus preserving it. In the present computer implementation, region synchronization is implemented by the propagation of a “phase index” as expected in a biological neural network. In the sky, the artifact (likely a tree limb) has disappeared, because one non-local effect of this process is small enough so such regions are absorbed though the non-linear diffusion. In a nutshell, we have here a “full resolution”, but simplified image.

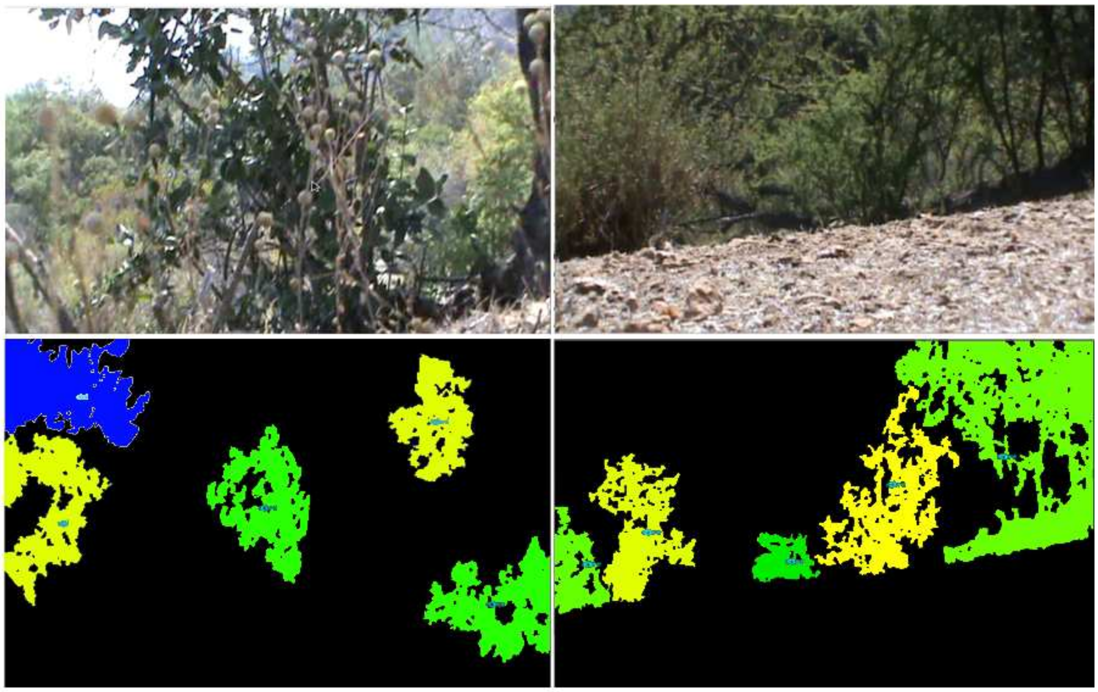

Figure 4: Two examples of detection of visual events (here, spatial events only). Top views correspond to the input image, bottom views to the detection output. The most salience regions are shown in both cases, i.e. those with maximal size and/or channel average values (here, for the present figures, size only is considered). *Left view* A "sky" zone and four dark/light "tree" zone have been detected. *Right view* The main "tree" zone have been detected, while the ground is less salient since not structured in terms of "regions". Clearly this kind of information is relevant to directly detect zones with potential pray/predators "hidden in the green". The categorization algorithm has been trained with a very small learning set (a few dozen of samples) and the recognition test is no more than a comparison of the region synchronized parameters (color, spectral signature) with respect to prototypes in competition. Such functionality is easy to generalize to another visual events (e.g. snake crawling, predator approaching motion, etc..) and thus provides an effective early-vision generation of a-priory information. We have numerically verified that such detection is robust along the image sequence. However, we also experimented that the non-linear filtering (i.e., segmentation) step is mandatory: Without this sophisticated early-vision step, the present categorization has very poor performances. This output represents sparse responses of the retina in the presence of specific visual events.